phenopix: a R package for image-based vegetation phenology

Gianluca Filippa\textsuperscript{a,}\textsuperscript{*}, Edoardo Cremonese\textsuperscript{a}, Mirco Migliavacca\textsuperscript{b}, Marta Galvagno\textsuperscript{a}, Matthias Forkel\textsuperscript{b}, Lisa Wingate\textsuperscript{c}, Enrico Tomelleri\textsuperscript{d}, Umberto Morra di Cella\textsuperscript{a}, Andrew D. Richardson\textsuperscript{e}

\textsuperscript{a}Environmental Protection Agency of Aosta Valley, ARPA Valle d’Aosta, Climate Change Unit (Italy)
\textsuperscript{b}Max Planck Institute for Biogeochemistry, Department Biogeochemical Integration (Jena, Germany)
\textsuperscript{c}INRA, UMR ISPA (Bordeaux, France)
\textsuperscript{d}Institute for Applied Remote Sensing, EURAC (Bolzano, Italy)
\textsuperscript{e}Harvard University, Department of Organismic and Evolutionary Biology (Cambridge, MA, USA)

Abstract

In this paper we extensively describe new software available as a R package that allows for the extraction of phenological information from time-lapse digital photography of vegetation cover. The \textit{phenopix} R package includes all steps in data processing. It enables the user to: draw a region of interest (ROI) on an image; extract red green and blue digital numbers (DN) from a seasonal series of images; depict greenness index trajectories; fit a curve to the seasonal trajectories; extract relevant phenological thresholds (phenophases); extract phenophase uncertainties.

The software capabilities are illustrated by analyzing one year of data from a selection of seven sites belonging to the PhenoCam network (\texttt{phenocam.sr.unh.edu/}), including an unmanaged subalpine grassland, a tropical grassland, a deciduous needle-leaf forest, three deciduous broad-leaf temperate forests and an evergreen needle-leaf forest. One of the novelties introduced by the package is the spatially explicit, pixel-based analysis, which potentially allows to extract within-ecosystem or within-individual variability of phenology. We examine the rela-

\textsuperscript{*}Corresponding author
Email address: g.filippa@arpa.vda.it (Gianluca Filippa)
tionship between phenophases extracted by the traditional ROI-averaged and
the novel pixel-based approaches, and further illustrate potential applications
of pixel-based image analysis available in the phenopix R package.

Keywords:  image analysis, community ecology, pixel-based analysis, phenology

1. Introduction

Traditional monitoring of plant phenology relies on direct human observations of discrete phenological events, or phenophases, such as bud-burst, flowering, autumn decoloring, and leaf-fall (e.g. Lechowicz 1984; Richardson et al. 2006; Galvagno et al. 2013; Migliavacca et al. 2008; Filippa et al. 2015). Such observations are typically made on a limited number of individual organisms, across a limited geographic area (i.e., often at a specific research site). On the other hand, satellite remote sensing allows observing land surface phenology on regional to global scales but has a limited representativeness for phenological changes at ecosystem or species-level (White and Nemani 2006; Delbart et al. 2005; Busetto et al. 2010; Hufkens et al. 2012; Forkel et al. 2015). At an intermediate scale, near-surface remote sensing of phenology makes use of radiometric instruments or imaging sensors. Near-surface remote sensing quantifies, at high temporal resolution, and with a flexible degree of spatial integration (i.e., the potential to look across the canopy as a whole, but at the same time focus on individual organisms), seasonal changes in the optical properties of vegetated surfaces (e.g. Jenkins et al. 2007; Richardson et al. 2007; Soudani et al. 2012). Recent studies have demonstrated the soundness of digital cameras as multi-channel imaging sensors (Richardson et al. 2009; Klosterman et al. 2014; Wingate et al. 2015; Migliavacca et al. 2011).

In the recent literature, different approaches have been used to process digital images of the vegetation canopy. Several researchers have exhaustively addressed the issues of data quality and data filtering in order to reduce noise in the seasonal trajectories of greenness (e.g. Sonntag et al. 2012; Julitta
et al., 2014; Migliavacca et al., 2011). Other authors focused on curve fitting/smoothing methods for extracting dates from phenological time-series (i.e. phenophases, see e.g. Zhang et al. 2003; Beck et al. 2006; Gu et al. 2009; Elmore et al. 2012; Klosterman et al. 2014 and the greenbrown R package). The full exploitation of the existing plethora of methods and approaches would benefit from a comprehensive framework in which to compare them across ecosystem types and climate conditions.

Phenological networks based on the analysis of digital images are growing worldwide. In the US, PhenoCam network includes over 200 sites that use a standard camera and follow a standard protocol (phenocam.sr.unh.edu/webcam/). In Europe, the EUROPHEN network consists of about 60 flux sites equipped with digital cameras (Wingate et al., 2015). Similar networks are also present in Australia (www.aceas.org.au, Brown et al. in preparation) and Asia (Nasahara and Nagai, 2015). The increasing use of repeated digital photography for phenological research highlights the need of easy-to-use, open source and flexible software tools for the processing of the images. Additionally, the existing networks are poorly coordinated in terms of camera types, settings and sampling protocols, thus presenting a big challenge to build a flexible tool capable of facing the large diversity of ecosystems, image quality and setups.

In this paper we present a collection of functions packed in a software available as a R package (R Core Team, 2015), called phenopix (r-forge.r-project.org/projects/phenopix/).

We will first show the main features of the package and then illustrate its application on a selection of sites belonging to the PhenoCam dataset. Lastly, a section focused on pixel-based analysis will: 1) examine the relationship between phenological thresholds (phenophases) extracted from the average seasonal trajectory of greenness over a region of interest (ROI-averaged approach) and from each pixel of such region (pixel-based approach); 2) illustrate potential applications of pixel-based image analysis to discriminate between subtly different phenological seasonal trajectories within the same image scene.
2. PhenoCam sites

The study sites used to illustrate the functionality of the phenopix package belong to the PhenoCam network (phenocam.sr.unh.edu/webcam/) and are illustrated in table 1 and fig. 1.

<table>
<thead>
<tr>
<th>Site ID</th>
<th>Coords (deg)</th>
<th>Elev (m)</th>
<th>PFT</th>
<th>Camera type</th>
<th>Dominant species</th>
</tr>
</thead>
<tbody>
<tr>
<td>torgnon-ND</td>
<td>45.8N 7.6E</td>
<td>2160</td>
<td>GRA</td>
<td>Nikon D5000</td>
<td>Nardus stricta</td>
</tr>
<tr>
<td>kamuela</td>
<td>20.0N -155.7E</td>
<td>850</td>
<td>GRA</td>
<td>Stardot</td>
<td>C3/C4 grasses</td>
</tr>
<tr>
<td>bartlett</td>
<td>44.1N -71.3E</td>
<td>268</td>
<td>DBF</td>
<td>Axis 211</td>
<td>Acer rubrum; Fagus grandifolia</td>
</tr>
<tr>
<td>harvard</td>
<td>42.5N -72.1E</td>
<td>340</td>
<td>DBF</td>
<td>Stardot</td>
<td>Quercus rubra; Acer rubrum</td>
</tr>
<tr>
<td>mammothcave</td>
<td>37.2N -86.1E</td>
<td>226</td>
<td>DBF</td>
<td>Stardot</td>
<td>Quercus sp.; Carya sp.</td>
</tr>
<tr>
<td>torgnon-LD</td>
<td>45.8N 7.6E</td>
<td>2091</td>
<td>DNF</td>
<td>Nikon D5000</td>
<td>Larix decidua</td>
</tr>
<tr>
<td>harvardhemlock</td>
<td>42.5N -72.2E</td>
<td>345</td>
<td>ENF</td>
<td>Stardot</td>
<td>Tsuga canadensis</td>
</tr>
</tbody>
</table>

3. Main functions

The typical work-flow of the phenopix package is summarized in the flowchart shown in fig. 2. First, one (or more) region(s) of interest (ROI) is (are) chosen, then digital color numbers are extracted from the ROI of each image, and processed to obtain a seasonal time series. After filtering the time series, data are fitted with either a double logistic equation or a smoothing curve, on which phenological thresholds (phenophases) are extracted. Finally, uncertainty of the fit and of phenophases can be computed.

3.1. Regions of interest (ROIs)

The scene of the picture rarely includes only the targeted vegetation canopy, thus the user will want to choose a particular region within the scene for analysis. Even more often, more than one region may be of interest, for example in a
mixed forest one might independently analyze different deciduous species and evergreen trees (e.g. Ahrends et al., 2008). The function DrawROI() allows the user to draw one or more regions of interest on-screen, using the mouse cursor on a chosen reference picture.

3.2. Extract vegetation indices

From the digital color values of each image the green chromatic coordinate ($G_{CC}$) is computed. $G_{CC}$ is a vegetation index derived from photographic image and quantifies the greenness relative to the total brightness. $G_{CC}$ is computed as follows:

$$G_{CC} = \frac{G_{DN}}{R_{DN} + G_{DN} + B_{DN}}$$  \hspace{1cm} (1)

where $G_{DN}$, $R_{DN}$, $B_{DN}$ are the green, red and blue digital numbers, respectively (Gillespie et al., 1987). Similarly, chromatic coordinates of red and blue ($R_{CC}$ and $B_{CC}$) are also computed. Several indices based on RGB colors have been developed in the last years, including for example the green excess index (GEI) (Woebbecke et al., 1995; Mizunuma et al., 2013). Some authors also used a combination of $G_{CC}$ and $R_{CC}$ to extract autumn phenophases (e.g. Klosterman et al., 2014). For simplicity, all subsequent analyses will be focused on $G_{CC}$ but phenopix allows the analysis of the whole variety of color indices, including the computation of new ones.

The function extractVIs() extracts raw red, green and blue digital numbers from each pixel in the ROI, and computes color chromatic coordinates as in eq.1. Vegetation indices can be computed on ROIs with 2 approaches (fig. 2): 1) the ROI-averaged approach: color chromatic coordinates are computed for each pixel and then averaged over the whole ROI, and (2) the pixel-based approach, where each pixel belonging to the ROI is analyzed separately (section 5). The procedure is repeated for each image in the archive. A time-stamp is retrieved from the file name of the image and a time series of the computed indices is returned. Specific rules must be followed in naming the image files,
and the reader is referred to the package help pages for more details.

3.3. Data filtering

Data retrieved from images often need robust methods for filtering the time series. Bad weather conditions, low illumination and dirty lenses are among the most common issues that determine noise in the time series of vegetation indices. Previous studies (e.g. Sonnentag et al., 2012; Julitta et al., 2014; Migliavacca et al., 2011; Papale et al., 2006) provide the background upon which the filtering techniques implemented in phenopix were chosen. We designed a function `autoFilter()` based on 5 different approaches.

(1) A filter called `night` which removes $G_{CC}$ records lower than a certain threshold, by default 0.2. This filters removes night data or images with very scarce illumination (e.g. early morning or sunset images, cloudy days, etc.).

(2) A filter called `blue`, based on blue chromatic coordinate ($B_{CC}$) (Julitta et al., 2014). First, hourly $B_{CC}$ data is aggregated at a daily resolution and daily means and standard deviations are computed. A quantile (by default the 5th percentile) is then calculated on standard deviations. This threshold is then added and subtracted from the daily mean $B_{CC}$ to generate a seasonal envelope. Raw data falling outside this envelope are discarded. The `blue` filter was designed to remove images with dominant clouds and/or snow on the canopy, because the blue chromatic coordinate was found to be very sensitive to such conditions.

(3) A filter called `mad`, following the method of Papale et al. (2006), an outlier detection based on the double-differenced time series, using the median of absolute deviation about the median (MAD) that is a robust outlier estimator.

(4) A filter called `spline`, following the method of Migliavacca et al. (2011), based on recursive spline smoothing and residual computation followed by removal of outliers falling outside a given residual envelope.

(5) A filter called `max`, following the method of Sonnentag et al. (2012), based on the identification of the 90th percentile values in three-days moving windows.
Filters can be applied alone or in sequence so that the user can choose the best combination of filters that suits the image archive to be processed. Additionally, each filter has its own discarding criteria that may be tuned by the user according to the quality of the input G\textsubscript{CC} time series. The default behavior of the filtering function is to use a sequence of \textit{night}, \textit{spline} and \textit{max} filters. In general, this sequence will be effective enough to properly filter the G\textsubscript{CC} time series. The user is advised to apply the \textit{max} filter to effectively minimize the impact of changes in scene illumination \citep{Sonnentag2012}.

### 3.4. Fit a curve to the G\textsubscript{CC} seasonal course and extract phenophases

The extraction of phenophases is done in two steps \citep{Forkel2015,Klosterman2014}. First, a curve is fitted to the G\textsubscript{CC} seasonal curve to reduce the influence of single observation and to better capture the seasonal behaviour. In a second step different extraction methods can be used to retrieve phenophase metrics from the fitted seasonal curve. We selected four different double logistic equations to be included in the package (table 2). The equations differ in the number of parameters to be optimized and hence in the flexibility of the fitting curves. For a thorough examination of the fittings and explanation of curve parameters, see the correspondent publications. In addition to the double logistic equations, an approach based on a smoothed cubic spline is also available. There are cases, e.g. with very noisy time series or weak signal (low seasonal amplitude in greenness) where the double logistic fits fail. In such cases, the spline method is the only possibility to extract a seasonal trajectory.

<table>
<thead>
<tr>
<th>Equation</th>
<th>Reference</th>
<th>Option name</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(t) = m_{\text{in}} + (m_{\text{max}} - m_{\text{in}}) \cdot \frac{1 + e^{-\left(\frac{t - \text{r}<em>{\text{start}}}{\text{s}</em>{\text{start}}}\right)}}{1 + e^{-\left(\frac{t - \text{s}<em>{\text{start}}}{\text{r}</em>{\text{start}}}\right)}} )</td>
<td>Beck et al. \citeyear{Beck2006}</td>
<td>&quot;beck&quot;</td>
</tr>
<tr>
<td>( f(t) = m_{\text{1}} + \left(m_{\text{2}} - m_{\text{1}}\right) \cdot \frac{1 + e^{-\left(\frac{t - \text{r}<em>{\text{start}}}{\text{s}</em>{\text{start}}}\right)}}{1 + e^{-\left(\frac{t - \text{s}<em>{\text{start}}}{\text{r}</em>{\text{start}}}\right)}} )</td>
<td>Elmore et al. \citeyear{Elmore2012}</td>
<td>&quot;elmore&quot;</td>
</tr>
<tr>
<td>( f(t) = (a_{1}t + b_{1}) + (a_{2}t^{2} + b_{2}t + c) \cdot \frac{1 + e^{-\left(\frac{t - \text{r}<em>{\text{start}}}{\text{s}</em>{\text{start}}}\right)}}{1 + e^{-\left(\frac{t - \text{s}<em>{\text{start}}}{\text{r}</em>{\text{start}}}\right)}} )</td>
<td>Klosterman et al. \citeyear{Klosterman2014}</td>
<td>&quot;klosterman&quot;</td>
</tr>
<tr>
<td>( f(t) = y_{\text{0}} + \left(\frac{a_{1}}{1 + e^{-\left(\frac{t - \text{r}<em>{\text{start}}}{\text{s}</em>{\text{start}}}\right)}}\right) - \left(\frac{a_{2}}{1 + e^{-\left(\frac{t - \text{r}<em>{\text{stop}}}{\text{s}</em>{\text{stop}}}\right)}}\right) )</td>
<td>Gu et al. \citeyear{Gu2009}</td>
<td>&quot;gs&quot;</td>
</tr>
</tbody>
</table>
After curve fitting, a suite of different approaches is available to extract phenophases, as summarized in table 3 and illustrated in figure 3. The phenophase extraction methods lead to the definition of dates that may have markedly different ecological meanings and be useful for several applications in the field of environmental sciences. However, the discussion of such meanings is beyond the objectives of this presentation paper, and the reader is referred to the publications listed in table 2 for further information. Here it is important to point out that the availability of very simple methods (e.g. trs) and more sophisticated ones (e.g. Klosterman method) allow the user to choose the extraction method better suited to the phenological trajectory of the investigated ecosystem and/or based on the quality of the input data.

Table 3: Methods for phenophase extraction available in phenopix. sos: start of season, eos: end of season, los: length of season, pop: peak of season position, peak: maximum seasonal $G_{CC}$, mgs: mean growing season $G_{CC}$, msp: mean spring $G_{CC}$, mau: mean autumn $G_{CC}$, rsp: rate of spring greenup, rau: rate of autumn senescence, UD: upturn date, SD: stabilization date, DD: downturn date, RD: recession date, prr: peak recovery rate, psr: peak senescence rate

<table>
<thead>
<tr>
<th>Name</th>
<th>Phenophases</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>trs</td>
<td>sos, eos, los, pop, mgs, peak, msp, mau</td>
<td>Based on a user-defined threshold of seasonal development of $G_{CC}$</td>
</tr>
<tr>
<td>derivatives</td>
<td>as TRS plus rsp and rau</td>
<td>Based on local extremes in the first derivative</td>
</tr>
<tr>
<td>klosterman</td>
<td>Greenup, Maturity, Senescence, Dormancy</td>
<td>Based on local extremes in the rate of change of curvature k (Kline, 1998)</td>
</tr>
<tr>
<td>gu</td>
<td>UD, SD, DD, RD, prr, psr</td>
<td>Based on a combination of local maxima in the first derivative (Gu et al., 2009)</td>
</tr>
</tbody>
</table>

The combination of four fitting methods plus the cubic spline smoothing
with the four phenophase extraction methods produces 20 different available combinations in output from a single yearly $G_{CC}$ time series. A specific function $\text{greenExplore()}$ is designed to give an overview of all fittings and phases for a given dataset (fig. 4). The RMSE for each fit is also shown, so that the user can use it as a criterion for fit selection. Once the fit is chosen, by examining the plot along a row the user can evaluate which phenophase method is more suitable to the $G_{CC}$ time series. The possibility to combine different fitting and phenophase methods to a $G_{CC}$ time series provides a framework in which to compare in detail the different methods currently in use. This need was highlighted by Keenan et al. (2014) in a recent attempt to link forest phenology as described by time-lapse photography and physiology as described by traditional plant trait measurements.

In addition to the above described methods, $\text{phenopix}$ also implements a phenophase extraction method (function $\text{PhenoBP()}$) based on linear piecewise regression and correspondent break points in the time series (Wingate et al., 2015). This method was designed to accommodate multiple greening peaks during the same season, which is typical, for example, of water limited ecosystems such as Kamuela (fig 5) or managed grasslands and croplands.

4. Estimation of the uncertainty

Traditionally, the analysis of digital images for phenology has rarely included the estimation of uncertainty on phenophase extraction, despite its paramount importance (e.g. for the evaluation of method robustness and as input data for the optimization of process-based or land surface models, where phenology is currently poorly represented (Richardson et al., 2013)).

The $\text{phenopix}$ package provides two approaches for the estimation of uncertainty, one for the fitted double logistic equations and one for the smoothing spline method. For the fitted equations the residuals between the model fit and the observed data are used to generate random noise and fitting is applied recursively to randomly-noised original data. This procedure results in
an ensemble of curves and curve parameters on which phenophase extraction is performed, thus providing an uncertainty estimate of phenophases. For the spline smoothing method, the uncertainty is estimated by combining a random noise as for the fitted equations with multiple changes in the spline’s degrees of freedom, to account for the arbitrariness in their choice. In particular, spline degrees of freedom are set at 5% of the time series length and then let vary recursively between 1% and 5% of the time series length to generate the ensemble of smoothing curves on which the uncertainty is estimated. In addition to the residuals method, for the fitted equations only, we are also developing a method based on the hessian matrix of the curve parameters.

Figure 6 shows the uncertainty analysis for Harvard deciduous 2013 GCC seasonal trajectory, with 500 replicates. The fitting method applied in this case was Klosterman. By default, the confidence interval is represented by the 10th and 90th percentile of the distribution of all extracted phenophases, but other options are available. In any case, the user can access the uncertainty data tables of (1) curve equation parameters and (2) extracted phenophases to customize the computation of the uncertainty envelope.

5. Phenopix application: spatial analysis

Based on an analysis run on all sites included in this paper except for Kamuela (i.e. six year-sites), we present in this paragraph the application of the pixel based analysis to investigate the spatial variability of phenology within a ROI (pixel-based approach).

To date, few studies have explored the possibility of analyzing a set of images pixel by pixel rather than averaging color values on the overall region of interest ([Julitta et al. 2014, Ide and Oguma 2013]). Therefore, an emergent question is how the ROI-averaged analysis is representative of all ROI pixels in terms of extracted phenophases? And, is this relationship consistent across fitting methods and ecosystem types? The phenopix package allows the user to perform pixel based image analysis and provides a number of functions to display and
analyze the results. It must be noted that pixel-by-pixel analysis is computationally intense. We report in Table 4 the computation times required to analyze a ROI containing 10000 pixels on a seasonal time series of 5000 images with all fits available in phenopix for 3 different computers, using one single processor. Computation time may be as high as 70 hours, with very low values for spline smoothing and higher for data filtering and fitted equations. However, parallel computation available in the spatial functions of phenopix enable considerably reduced computation times compared to those reported in Table 4.

Table 4: Estimated computation time (hours) required to complete different steps of the pixel-based analysis processing 10000 pixels from a seasonal time series of 5000 images using one single processor on three different computers

<table>
<thead>
<tr>
<th>step</th>
<th>3.07 GHz</th>
<th>2.60 GHz</th>
<th>2.13 GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>filtering</td>
<td>28</td>
<td>49</td>
<td>52</td>
</tr>
<tr>
<td>spline</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>beck</td>
<td>36</td>
<td>63</td>
<td>69</td>
</tr>
<tr>
<td>elmore</td>
<td>0.7</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>klosterman</td>
<td>10</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>gu</td>
<td>22</td>
<td>37</td>
<td>42</td>
</tr>
</tbody>
</table>

We evaluate the relationship between approaches (i.e. ROI-average vs pixel-based) and fitting/phenophase methods across sites in Fig. 7. Two different fitting methods (namely Klosterman and spline) applied to each pixel of the ROIs and then averaged are in good agreement with each other (Fig. 7, first above diagonal panel), with a slightly higher correlation for spring than for autumn phases. The error bars for this relationship suggest a higher variability for autumn than for spring phenophases (first below diagonal panel). When comparing the ROI-averaged and the pixel-based approach, the best relationship is found with the Klosterman fitting method applied to both approaches, with correlation coefficients of 0.97 and 0.75 for spring and autumn phenology, respectively. However, the relationship between spline fit applied pixel-by-pixel
and Klosterman fit applied to the average ROI also shows very good agreement. In contrast, the spline method applied to the average ROI (last column in fig. 7) has a consistently weaker relationship with all other methods, resulting in earlier autumn phases and a much higher variability.

Across phenophase methods (different symbols in fig. 7), there is no evidence of a systematic bias between the two approaches for any particular phase extraction method. Across ecosystem types, needle-leaf forest (ENF, i.e. harvardhemlock site) shows a consistent departure from the 1:1 line even in the best relationships (third panel from left in the first row of fig. 7), probably due to the lower signal to noise ratio in the seasonal trajectory of evergreen trees. However, the most striking difference is the consistently lower correlation for autumn than for spring phases, suggesting that in the same ecosystem and possibly also within a single species autumn phases are more variable than spring phases.

In summary, the faster spline smoothing applied pixel-by-pixel leads to the identification of phenophases in substantial agreement with the more computationally-intense Klosterman fit. Additionally, being more flexible, the spline fitting leads to a much lower number of unfitted pixels compared to the Klosterman fit. Hence, from this analysis, spline smoothing is preferred over curve fitting for pixel-by-pixel analysis. For the ROI-averaged approach, it is strongly preferable to perform curve fitting over spline smoothing, provided that the \( G_{CC} \) seasonal trajectory does not show multiple peaks, related to e.g. water stress or management practices.

One interesting application of the pixel-based analysis stems from the examination of the frequency distribution of a given phenophase across all pixels. When a phenophase shows a bi- (or multi-) modality and a consistent spatial distribution, pixels may be grouped according to the values of one (or more) phenophase(s) and separate seasonal trajectories can be obtained (fig. 8). By applying this procedure to the torgnon-ND grassland site we were able to identify a bimodal distribution in maturity phase (i.e. the end of spring growth), as extracted after Klosterman curve fitting. We then selected pixels with maturity onset dates falling around the two modes (i.e. DOY 180±3 and DOY 205±3)
and computed an average trajectory for these two groups. This was possible because in the pixel-based approach each pixel has associated curve parameters to reconstruct the seasonal trajectory. The resulting average trajectories are markedly distinct around the seasonal peak, and reflect the different ecology of forbs and grasses occurring over very short distances (<20 cm) at this grassland site (Julitta et al., 2014). Further interpretation of this behavior is beyond the objective of this paper, but this preliminary analysis highlights the potential use of spatially explicit image processing to identify different phenology related to different plant functional types.

More sophisticated clustering methods can be applied to pixel values in order to distinguish different phenological patterns. For example, at the Bartlett site (a mixed deciduous forest) we might expect a different phenology in early or late flushing species (fig. 9a). With a cluster analysis (Hartigan and Wong, 1979) using phenophases extracted with all available methods as an input matrix, we were able to clearly distinguish two portions of the canopy (fig. 9b). In this example the cluster analysis was used to define two sub-ROIs, that reentered the processing chain (ROI-averaged approach, fig. 2) and led to the identification of two markedly different seasonal trajectories (fig. 9c,d). Beech-dominated portions of the canopy (cluster 2) show a later yellowing/browning compared to maple, prevailing in cluster 1 (Richardson et al., 2009).

6. Summary and outlook

We presented a new R package called phenopix that collects the most up-to-date processing techniques for repeated digital photography of vegetation canopy in the context of phenological and ecological research. Together with the basic and well established processing steps, the package implements several new features, including the possibility to combine different fitting and phenophase methods, to compute uncertainty on extracted phenophases, and to conduct pixel-by-pixel analysis. Phenophase maps obtained by the pixel-based approach
may represent a promising tool for a variety of ecological analyses, including for example: 1) the study of species-specific phenology; 2) the examination of spatial variability of apparently homogeneous ecosystems such as grasslands; 3) the potential to discriminate between overstory and understory or age-related differences in phenological patterns of open canopy forests.

One major future step in the evolution of phenopix package includes the computation of the so called camera-NDVI, in the formulation proposed by [Petach et al. (2014)].

The phenopix R package offers a suite of standardized and reproducible processing code that may be adopted, deployed and further developed by the existing phenocameras networks worldwide.

Acknowledgments

ADR acknowledges support from the National Science Foundation, through the Macrosystems Biology program (grant EF-1065029).

The package was developed in the framework of the ALCOTRA Interreg Project n 227 e-PHENO Reti fenologiche nelle Alpi.

We thank Morgan Furze, Harvard University, for assistance with editing the manuscript.

References


Figure 1: A sample image from all PhenoCam sites included in this study masked on the chosen region of interest (ROI), along with the seasonal trajectories of filtered $G_{CC}$ in year 2013. Ecosystem type abbreviations are as in table 1.
Figure 2: The work-flow of the processing chain in the phenopix R package. Red and blue objects denote specific features of pixel-based and ROI-averaged approaches, respectively. Pixel-based approach produces a phenophase map (left), whereas ROI-averaged approach results in a GCC seasonal trajectory and a certain number of phenophases, depending on the method chosen (right). The example is from torgnon-ND grassland site in year 2013.
Figure 3: Illustration of the methods used to extract phenological thresholds (phenophases) in a seasonal $G_{CC}$ trajectory ($f(t)$). 

a) 

trs method: Phenophases are defined as the day of year (DOY) when the 50% $G_{CC}$ is reached either during greenup (sos) and autumn (eos), the DOY of maximum $G_{CC}$ is called pop; 

b) 

derivatives method: Phenophases are defined as the DOY when $f'(t)$ shows the absolute maximum and minimum;

c) 
Klosterman method: Phenophases are defined as the two local maxima (greenup) and two local minima (autumn) in the rate of change in curvature $k'$ (Kline, 1998; Klosterman et al., 2014) and are named after Zhang et al. (2003); 

d) 
Gu method: Maximum and minimum of $f'(t)$ are used to define slopes of recovery and senescence lines tangent to the curves (green and brown lines). The intersection between these lines and baseline and maxline define the four phenophases in the original formulation (Gu et al., 2009). To account for the mid season decrease in $G_{CC}$ we have further defined a plateau line as a linear fit to $G_{CC}$ values between SD and DD, in order to adjust the definition of phenophase DD. For phenophase abbreviations and details see table 3.
Figure 4: All fitting and phenophase methods applied to Harvard deciduous forest $G_{CC}$ seasonal trajectory in year 2013. This plot is the output from the functions `greenExplore()` and `plotExplore()`. Plots in the same row share the same fitting method, whereas those in the same column share the same phenophase extraction method. The RMSE for each fitting is also annotated in the first column.
Figure 5: Phenophases as determined by the break point analysis (PhenoBP() function) applied to Kamuela grassland $G_{GC}$ seasonal trajectory in year 2013. The thickness of the vertical dashed lines is proportional to the uncertainty of the extracted break points (same scale as the x-axis). Uncertainty is estimated using the function confint() as the 95% confidence interval.
Figure 6: The uncertainty analysis applied to Harvard deciduous forest $G_{CC}$ seasonal trajectory in year 2013. The grey lines represent the 500 simulated curves on which phenophases are extracted. Shaded areas on phenophases represent $10^{th}$ and $90^{th}$ percentiles of the 500 replications. Data are fitted using the Klosterman method. Phenophases are extracted with a) the trs method; b) the derivatives method, c) klosterman method, d) gu method.
Figure 7: Scatterplot matrix showing: a) above the diagonal, the relationship between phenophases extracted with two different approaches (pixel-based and ROI-averaged) and two different fitting methods (spline smoothing and Klosterman fit), and b) below the diagonal, errors associated. Error bars for pixel approach represent the mean absolute deviation of all pixels, whereas error bars for ROI-averaged approach represent the uncertainty as described in section 4. Note that the scale for error bars is provided in the first below-diagonal panel. Different symbols denote different phenophase extraction methods. Close symbols are for spring phenology and open symbols for autumn phenology. Different colors represent different PFTs (abbreviations for PFTs are as in table 1). Data from all sites except for Kamuela were used in this analysis.
Figure 8: Results of the spatial analysis conducted on torognon-ND grassland site. a) Map of the spatial distribution of maturity (Klosterman fitting method, Klosterman phenophases method), with darker colors indicating earlier occurrence of the phase (in DOYs). b) Average seasonal trajectory of pixels clustered according to the bimodal distribution of maturity. Patches with forbs dominating show an earlier maturity compared to grasses (dashed lines). In the inset, density distribution of Maturity phase across all pixels. Shaded areas denote the maturity intervals used to subset ROI pixels. The resulting subsets were in turn used to extract the seasonal trajectories of forbs- and grass-dominated portions of the ROI.
Figure 9: Results of the cluster analysis run on Bartlett pixel-based phenophases. a) Sample image from DOY 275, b) Distribution of clusters in the ROI, c) Density distribution of Greenup and Dormancy phases across all pixels, and d) Averaged seasonal trajectories for the two clusters. All extracted phases (from the four methods presented in section 3) were used in input to clustering.